

**Sequence Information:****SEQ ID NO: 1 (INSP005A nucleotide sequence exon 1)**

5       1 ATGGGTGGTA GTGGTGTCTG GGAGGTCCCC TTCCTGCTCT CCAGCAAGTA  
      51 CG

**SEQ ID NO: 2 (INSP005A protein sequence exon 1)**

10       1 MGGSGVVEVP FLLSSKYD

**SEQ ID NO: 3 (INSP005A nucleotide sequence exon 2)**

      1 ATGAGCCCAG CCGCCAGGTC ATCCTGGAGG CTCTTGCGGA GTTTGAACGT  
15       51 TCCACGTGCA TCAGGTTTGT CACCTATCAG GACCAGAGAG ACTTCATTTC  
      101 CATCATCCCC ATGTATGG

**SEQ ID NO: 4 (INSP005A protein sequence exon 2)**

20       1 EPSRQVILEA LAEFERSTCI RFVTYQDQRD FISIIPMYG

**SEQ ID NO: 5 (INSP005A nucleotide sequence exon 3)**

      1 GTGCTTCTCG AGTGTGGGGC GCAGTGGAGG GATGCAGGTG GTCTCCCTGG  
25       51 CGCCACGTG TCTCCAGAAG GGCCGGGGCA TTGTCCTTCA TGAGCTCATG  
      101 CATGTGCTGG GCTTCTGGCA CGAGCACACG CGGGCCGACC GGGACCGCTA  
      151 TATCCGTGTC AACTGGAACG AGATCCTGCC AG

**30 SEQ ID NO: 6 (INSP005A protein sequence exon 3)**

      1 CFSSVGRSGG MQVVS LAPTC LQKGRGIVLH ELMHVLGFWH EHTRADRDRY  
      51 IRVNWNEILP G

**35 SEQ ID NO: 7 (INSP005A nucleotide sequence exon 4)**

      1 GCTTTGAAAT CAACTTCATC AAGTCTCAGA GCAGCAACAT GCTGACGCCC  
      51 TATGACTACT CCTCTGTGAT GCACTATGGG AG

**40 SEQ ID NO: 8 (INSP005A protein sequence exon 4)**

      1 FEINFIKSQS SNMLTPYDYS SVMHYGR

**SEQ ID NO: 9 (INSP005A nucleotide sequence exon 5)**

45       1 GCTCGCCTTC AGCCGGCGTG GGCTGCCCAC CATCACACCA CTTTGGGCCC  
      51 CCAGTGTCCA CATCGGCCAG CGATGGAACC TGAGTGCCTC GGACATCACC  
      101 CGGGTCCTCA AACTCTACGG CTGCAGCCCA AGTGGCCCCA GGCCCCGTGG  
50       151 GAGAG

**SEQ ID NO: 10 (INSP005A protein sequence exon 5)**

1 LAFSRRGLPT ITPLWAPSVH IGQRWNLSAS DITRVLKLYG CSPSGPRPRG

51 RG

**SEQ ID NO: 11 (INSP005A nucleotide sequence exon 6)**

1 GGTCCCATGC CCACAGCACT GGTAGGAGCC CCGCCCCGGC CTCCCTATCT

51 CTGCAGCGGC TTTTGGAGGC ACTGTCGGCG GAATCCAGGA GCGCCGACCC

101 CAGTGGTTCC AGTGCGGGAG GCCAGCCCGT TCCTGCAGGG CCTGGGGAGA

151 GCCCACATGG GTGGGAGTCC CCTGCCCTGA AAAAGCTCAG TGCAGAGGCC

201 TCGGCAAGGC AGCCTCAGAC CCTAGCTTCC TCCCCAAGAT CAAGGCCTGG

251 AGCAGGTGCC CCCGGTGTTC CTCAGGAGCA GTCCTGGCTG GCCGGAGTGT

301 CCACCAAGCC CACAGTCCCA TCTTCAGAAG CAGGAATCCA GCCAGTCCCT

351 GTCCAGGGAA GCCCAGCTCT GCCAGGGGGC TGTGTACCTA GAAATCATTT

401 CAAGGGGATG TCCGAAGAT

**SEQ ID NO: 12 (INSP005A protein sequence exon 6)**

1 SHAHSTGRSP APASLSLQRL LEALSAESRS PDPSGSSAGG QVPVAGPGES

51 PHGWESPALK KLSAEASARQ PQTLASSPRS RPGAGAPGVA QEQSWLAGVS

101 TKPTVPSSEA GIQVPVQGS PALPGGCVPR NHFKGMSED

**SEQ ID NO: 13 (INSP005A full nucleotide sequence)**

1 ATGGGTGGTA GTGGTGTCTG GGAGGTCCCC TTCCTGCTCT CCAGCAAGTA

51 CGATGAGCCC AGCCGCCAGG TCATCCTGGA GGCTCTTGCG GAGTTTGAAC

101 GTTCCACGTG CATCAGGTTT GTCACCTATC AGGACCAGAG AGACTTCATT

151 TCCATCATCC CCATGTATGG GTGCTTCTCG AGTGTGGGGC GCAGTGGAGG

201 GATGCAGGTG GTCTCCCTGG CGCCACGTG TCTCCAGAAG GGCCGGGGCA

251 TTGTCTTCA TGAGCTCATG CATGTGCTGG GCTTCTGGCA CGAGCACACG

301 CGGGCCGACC GGGACCGCTA TATCCGTGTC AACTGGAACG AGATCCTGCC

351 AGGCTTTGAA ATCAACTTCA TCAAGTCTCA GAGCAGCAAC ATGCTGACGC

401 CCTATGACTA CTCCTCTGTG ATGCACTATG GGAGGCTCGC CTTACGCCGG

451 CGTGGGCTGC CCACCATCAC ACCACTTTGG GCGCCAGTG TCCACATCGG

501 CCAGCGATGG AACCTGAGTG CCTCGGACAT CACCCGGGTC CTCAAACCTCT

551 ACGGCTGCAG CCCAAGTGGC CCCAGGCCCC GTGGGAGAGG GTCCCATGCC

601 CACAGCACTG GTAGGAGCCC CGCCCCGGCC TCCCTATCTC TGCAGCGGCT

651 TTTGGAGGCA CTGTCGGCGG AATCCAGGAG CCGCGACCCC AGTGGTTCCA

701 GTGCGGGAGG CCAGCCCGTT CCTGCAGGGC CTGGGGAGAG CCCACATGGG

751 TGGGAGTCCC CTGCCCTGAA AAAGCTCAGT GCAGAGGCCT CGGCAAGGCA  
801 GCCTCAGACC CTAGCTTCCT CCCCAAGATC AAGGCCTGGA GCAGGTGCCC  
5 851 CCGGTGTTGC TCAGGAGCAG TCCTGGCTGG CCGGAGTGTC CACCAAGCCC  
901 ACAGTCCCAT CTTCAGAAGC AGGAATCCAG CCAGTCCCTG TCCAGGGAAG  
951 CCCAGCTCTG CCAGGGGGCT GTGTACCTAG AAATCATTTT AAGGGGATGT  
10 1001 CCGAAGAT

**SEQ ID NO: 14 (INSP005A full protein sequence)**

1 MGGSGVVEVP FLLSSKYDEP SRQVILEALA EFERSTCIRF VTYQDQDFI  
15 51 SIIPMYGCFS SVGRSGGMQV VSLAPTCLQK GRGIVLHELM HVLGFWHEHT  
101 RADRDYIRV NWNELPGFE INFIKSQSSN MLTPYDYSSV MHYGRLAFSR  
20 151 RGLPTITPLW APSVHIGQRW NLSASDITRV LKLYGCSPSG PRPRGRGSHA  
201 HSTGRSPAPA SLSLQRLLEA LSAESRSPDP SGSSAGGQPV PAGPGESPHG  
251 WESPALKKLS AEASARQPQT LASSPRSRPG AGAPGVAQEQ SWLAGVSTKP  
25 301 TVPSSEAGIQ PVPVQGSPAL PGGCVPRNHF KGMSD

**SEQ ID NO: 15 (INSP005B nucleotide sequence exon 1)**

1 ATGGAGGGTG TAGGGGGTCT CTGGCCTTGG GTGCTGGGTC TGCTCTCCTT  
30 51 GCCAG

**SEQ ID NO: 16 (INSP005B protein sequence exon 1)**

1 MEGVGGLWPW VLGLLSLPG  
35

**SEQ ID NO: 17 (INSP005B nucleotide sequence exon 2)**

1 GTGTGATCCT AGGAGCGCCC CTGGCCTCCA GCTGCGCAGG AGCCTGTGGT  
51 ACCAGCTTCC CAGATGGCCT CACCCCTGAG GGAACCCAGG CCTCCGGGGA  
40 101 CAAGGACATT CCTGCAATTA ACCAAG

**SEQ ID NO: 18 (INSP005B protein sequence exon 2)**

1 VILGAPLASS CAGACGTSFP DGLTPEGTQA SGDKDIPAIN QG  
45

**SEQ ID NO: 19 (INSP005B nucleotide sequence exon 3)**

1 GGCTCATCCT GGAAGAAACC CCAGAGAGCA GCTTCCTCAT CGAGGGGGAC  
51 ATCATCCGGC CG  
50

**SEQ ID NO: 20 (INSP005B protein sequence exon 3)**

1 LILEETPESS FLIEGDIIRP

**SEQ ID NO: 21 (INSP005B nucleotide sequence exon 4)**

55 1 AGTCCCTTCC GACTGCTGTC AGCAACCAGC AACAAATGGC CCATGGGTGG

51 TAGTGGTGTC GTGGAGGTCC CCTTCCTGCT CTCCAGCAAG TACG

**SEQ ID NO: 22 (INSP005B protein sequence exon 4)**

5 1 SPFRLLSATS NKWPMGGSGV VEVFLLSSK YD

**SEQ ID NO: 23 (INSP005B nucleotide sequence exon 5)**

1 ATGAGCCCAG CCGCCAGGTC ATCCTGGAGG CTCTTGCGGA GTTTGAACGT  
10 51 TCCACGTGCA TCAGGTTTGT CACCTATCAG GACCAGAGAG ACTTCATTTT  
101 CATCATCCCC ATGTATGG

**SEQ ID NO: 24 (INSP005B protein sequence exon 5)**

15 1 EPSRQVILEA LAEFERSTCI RFVTYQDQRD FISIIPMYG

**SEQ ID NO: 25 (INSP005B nucleotide sequence exon 6)**

1 GTGCTTCTCG AGTGTGGGGC GCAGTGGAGG GATGCAGGTG GTCTCCCTGG  
20 51 CGCCACGTG TCTCCAGAAG GGCCGGGGCA TTGTCCTTCA TGAGCTCATG  
101 CATGTGCTGG GCTTCTGGCA CGAGCACACG CGGGCCGACC GGGACCGCTA  
151 TATCCGTGTC AACTGGAACG AGATCCTGCC AG  
25

**SEQ ID NO: 26 (INSP005B protein sequence exon 6)**

1 CFSSVGRSGG MQVVSLAPTC LQKGRGIVLH ELMHVLGFWH EHTRADRDRY  
30 51 IRVNWNEILP G

**SEQ ID NO: 27 (INSP005B nucleotide sequence exon 7)**

1 GCTTTGAAAT CAACTTCATC AAGTCTCGGA GCAGCAACAT GCTGACGCCC  
35 51 TATGACTACT CCTCTGTGAT GCACTATGGG AG

**SEQ ID NO: 28 (INSP005B protein sequence exon 7)**

1 FEINFIKSRS SNMLTPYDYS SVMHYGR

**SEQ ID NO: 29 (INSP005B nucleotide sequence exon 8)**

40 1 GCTCGCCTTC AGCCGGCGTG GGCTGCCCAC CATCACACCA CTTTGGGCCC  
51 CCAGTGTCCA CATCGGCCAG CGATGGAACC TGAGTGCCTC GGACATCACC  
101 CGGGTCCTCA AACTCTACGG CTGCAGCCCA AGTGGCCCCA GGCCCCGTGG  
45 151 GAGAG

**SEQ ID NO: 30 (INSP005B protein sequence exon 8)**

50 1 LAFSRRGLPT ITPLWAPSVH IGQRWNLSAS DITRVLKLYG CSPSGPRPRG  
51 RG

**SEQ ID NO: 31 (INSP005B nucleotide sequence exon 9)**

1 GGTCCCATGC CCACAGCACT GGTAGGAGCC CCGCTCCGGC CTCCCTATCT  
5 51 CTGCAGCGGC TTTTGGAGGC ACTGTCGGCG GAATCCAGGA GCCCCGACCC  
101 CAGTGGTTCC AGTGCGGGAG GCCAGCCCGT TCCTGCAGGG CCTGGGGAGA  
151 GCCCACATGG GTGGGAGTCC CCTGCCCTGA AAAAGCTCAG TGCAGAGGCC  
10 201 TCGGCAAGGC AGCCTCAGAC CCTAGCTTCC TCCCCAAGAT CAAGGCCTGG  
251 AGCAGGTGCC CCCGGTGTG CTCAGGAGCA GTCCTGGCTG GCCGGAGTGT  
301 CCACCAAGCC CACAGTCCCA TCTTCAGAAG CAGGAATCCA GCCAGTCCCT  
15 351 GTCCAGGGAA GCCCAGCTCT GCCAGGGGGC TGTGTACCTA GAAATCATTT  
401 CAAGGGGATG TCCGAAGAT

**20 SEQ ID NO: 32 (INSP005B protein sequence exon 9)**

1 SHAHSTGRSP APASLSLQRL LEALSAESRS PDPSGSSAGG QPVPAGPGES  
51 PHGWESPALK KLSAEASARQ PQTlassPRS RPGAGAPGVA QEQSWLAGVS  
25 101 TKPTVPSSEA GIQVPVQGS PALPGGCVPR NHFKGMSSED

**SEQ ID NO: 33 (INSP005B full nucleotide sequence)**

1 ATGGAGGGTG TAGGGGTCT CTGGCCTTGG GTGCTGGGTC TGCTCTCCTT  
30 51 GCCAGGTGTG ATCCTAGGAG CGCCCTGGC CTCCAGCTGC GCAGGAGCCT  
101 GTGGTACCAG CTTCCAGAT GGCCTCAGCC CTGAGGGAAC CCAGGCCTCC  
151 GGGGACAAGG ACATTCTCTG AATTAACCAA GGGCTCATCC TGAAGAAAC  
35 201 CCCAGAGAGC AGCTTCCTCA TCGAGGGGGA CATCATCCGG CCGAGTCCCT  
251 TCCGACTGCT GTCAGCAACC AGCAACAAAT GGCCCATGGG TGGTAGTGGT  
40 301 GTCGTGGAGG TCCCTTCCT GCTCTCCAGC AAGTACGATG AGCCAGCCG  
351 CCAGGTCATC CTGGAGGCTC TTGCGGAGTT TGAACGTTCC ACGTGCATCA  
401 GGT'TTGTAC CTATCAGGAC CAGAGAGACT TCATTTCCAT CATCCCCATG  
45 451 TATGGGTGCT TCTCGAGTGT GGGGCGCAGT GGAGGGATGC AGGTGGTCTC  
501 CCTGGCGCCC ACGTGTCTCC AGAAGGGCCG GGGCATTGTC CTTTATGAGC  
50 551 TCATGCATGT GCTGGGCTTC TGGCACGAGC ACACGCGGGC CGACCGGGAC  
601 CGCTATATCC GTGTCAACTG GAACGAGATC CTGCCAGGCT TTGAAATCAA  
651 CTTTATCAAG TCTCGGAGCA GCAACATGCT GACGCCCTAT GACTACTCCT  
55 701 CTGTGATGCA CTATGGGAGG CTCGCCTTCA GCCGGCGTGG GCTGCCCACC  
751 ATCACACCAC TTTGGGCCCC CAGTGTCCAC ATCGGCCAGC GATGGAACCT  
60 801 GAGTGCCTCG GACATCAGCC GGGTCTTCAA ACTCTACGGC TGCAGCCCAA  
851 GTGGCCCCAG GCCCCGTGGG AGAGGGTCCC ATGCCACAG CACTGGTAGG  
901 AGCCCCGCTC CGGCCTCCCT ATCTCTGCAG CGGCTTTTGG AGGCACTGTC

951 GGC GGAATCC AGGAGCCCCG ACCCCAGTGG TTCCAGTGCG GGAGGCCAGC  
 1001 CCGTTCCTGC AGGGCCTGGG GAGAGCCCAC ATGGGTGGGA GTCCCCTGCC  
 5 1051 CTGAAAAAGC TCAGTGCAGA GGCCTCGGCA AGGCAGCCTC AGACCCTAGC  
 1101 TTCTTCCCCA AGATCAAGGC CTGGAGCAGG TGCCCCCGGT GTTGCTCAGG  
 10 1151 AGCAGTCCTG GCTGGCCGGA GTGTCCACCA AGCCCACAGT CCCATCTTCA  
 1201 GAAGCAGGAA TCCAGCCAGT CCCTGTCCAG GGAAGCCCAG CTCTGCCAGG  
 1251 GGGCTGTGTA CCTAGAAATC ATTTCAAGGG GATGTCCGAA GAT  
 15

**SEQ ID NO: 34 (INSP005B full protein sequence)**

1 MEGVGGLWPV VLGLLSLPGV ILGAPLASSC AGACGTSFPD GLTPEGTQAS  
 51 GDKDIPAINQ GLILEETPES SFLIEGDIIR PSPFRLLSAT SNKWPMGGSG  
 20 101 VVEVPFLLSS KYDEPSRQVI LEALAEFERS TCIRFVTYQD QRD FISIIPM  
 151 YGCFSSVGRS GGMQVVSLAP TCLQKGRGIV LHELMHVLGF WHEHTRADRD  
 25 201 RYIRVNWNEI LPGFEINFIK SRSSNMLTPY DYSSVMHYGR LAFSRRGLPT  
 251 ITPLWAPSVH IGQRWNLSAS DITRVLKLYG CSPSGPRPRG RGS SHAHSTGR  
 30 301 SPAPASLSLQ RLLEALSAES RSPDPSGSSA GGQVPAPGPG ESPHGWESPA  
 351 LKKLSAEASA RQPQTLASSP RSRPGAGAPG VAQE QSWLAG VSTKPTVPSS  
 401 EAGIQPVVPVQ GSPALPGGCV PRNHFKGMSE D

**35 SEQ ID NO: 35 (INSP005bmature nucleotide sequence)**

1 GCGCCCCCTGG CCTCCAGCTG CGCAGGAGCC TGTGGTACCA GCTTCCCAGA  
 51 TGGCCTCACC CCTGAGGGAA CCCAGGCCTC CGGGGACAAG GACATTCTCTG  
 101 CAATTAACCA AGGGCTCATC CTGGAAGAAA CCCCAGAGAG CAGCTTCTCTC  
 151 ATCGAGGGGG ACATCATCCG GCCGAGTCCC TTCCGACTGC TGTCAGCAAC  
 40 201 CAGCAACAAA TGGCCCATGG GTGGTAGTGG TGTCGTGGAG GTCCCCTTCC  
 251 TGCTCTCCAG CAAGTACGAT GAGCCCAGCC GCCAGGTCAT CCTGGAGGCT  
 301 CTTGCGGAGT TTGAACGTT CACGTGCATC AGGTTTGTCA CCTATCAGGA  
 351 CCAGAGAGAC TTCATTTCCA TCATCCCCAT GTATGGGTGC TTCTCGAGTG  
 401 TGGGGCGCAG TGGAGGGATG CAGGTGGTCT CCCTGGCGCC CACGTGTCTC  
 45 451 CAGAAGGGCC GGGGCATTGT CCTTCATGAG CTCATGCATG TGCTGGGCTT  
 501 CTGGCACGAG CACACGCGGG CCGACCGGGA CCGCTATATC CGTGTCAACT  
 551 GGAACGAGAT CCTGCCAGGC TTTGAAATCA ACTTCATCAA GTCTCGGAGC  
 601 AGCAACATGC TGACGCCCTA TGA CTACTCC TCTGTGATGC ACTATGGGAG  
 651 GCTCGCCTTC AGCCGGCGTG GGCTGCCCAC CATCACACCA CTTTGGGCCC  
 50 701 CCAGTGTCCA CATCGGCCAG CGATGGAACC TGAGTGCCTC GGACATCACC  
 751 CGGGTCCTCA AACTCTACGG CTGCAGCCCA AGTGGCCCCA GGCCCCGTGG  
 801 GAGAGGGTCC CATGCCCACA GCACTGGTAG GAGCCCCGCT CCGGCCTCCC  
 851 TATCTCTGCA GCGGCTTTTG GAGGCACTGT CGGCGGAATC CAGGAGCCCC

901 GACCCCAAGTG GTTCCAGTGC GGGAGGCCAG CCCGTTCTCTG CAGGGCCTGG  
 951 GGAGAGCCCA CATGGGTGGG AGTCCCCTGC CCTGAAAAAG CTCAGTGCAG  
 1001 AGGCCTCGGC AAGGCAGCCT CAGACCCTAG CTTCTCTCCC AAGATCAAGG  
 1051 CCTGGAGCAG GTGCCCCCGG TGTTCGCTCAG GAGCAGTCCT GGCTGGCCGG  
 5 1101 AGTGTCCACC AAGCCCACAG TCCCATCTTC AGAAGCAGGA ATCCAGCCAG  
 1151 TCCCTGTCCA GGAAGCCCA GCTCTGCCAG GGGGCTGTGT ACCTAGAAAT  
 1201 CATTTCAAGG GGATGTCCGA AGAT

**SEQ ID NO: 36 (INSP005b mature polypeptide sequence)**

10 1 APLASSCAGA CGTSFPDGLT PEGTQASGDK DIPAINQLI LEETPESSFL  
 51 IEGDIIRPSP FRLLSATSNK WPMGGSGVVE VPFLSSKYD EPSRQVILEA  
 101 LAEFERSTCI RFVTYQDQRD FISIIPMYGC FSSVGRSGGM QVVS LAPTCL  
 151 QKGRGIVLHE LMHVLGFWHE HTRADRDRYI RVNWNELPG FEINFIKSR  
 201 SNMLTPYDYS SVMHYGRLAF SRRGLPTITP LWAPSVHIGQ RWNLSASDIT  
 15 251 RVLKLYGCSP SGPRPRGRGS HAHSTGRSPA PASLSLQRL EALSAESRSP  
 301 DPSGSSAGGQ PVPAGPGESP HGWESPALKK LSAEASARQP QTLASSPRSR  
 351 PGAGAPGVAQ EQSWLAGVST KPTVPSSEAG IQPVPVQGS ALPGGCVPRN  
 401 HFKGMSD

**20 SEQ ID NO: 37 (INSP005 Predicted Polypeptide Sequence)**

1 MLRLWDFNPG GALSDLALGL RMEEGGYSC AGACGTSFPD GLTPEGTQAS GDKDIPAINQ  
 61 GLILEETPES SFLIEGDIIR PSPFRLLSAT SNKWPMGGSG VVEVPFLSS KYDEPSHQVI  
 121 LEALAEFERS TCIRFVTYQD QRDFISIIPM YGCFSSVGRS GGMQVVS LAP TCLQKGRGIV  
 181 LHELMHVLGF WHEHTRADR RYIRVNWNEI LPGFEINFIK SQSSNMLTPY DYSSVMHYGR  
 25 241 LAFSRRGLPT ITPLWAPSVH IGQRWNLSAS DITRVLKLYG CSPSGPRPRG RGEWHGRKVT

**SEQ ID NO: 38 (pCR4 TOPO IPAAA78836-1 plasmid nucleotide sequence)**

1 AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCATTAA TGCAGCTGGC  
 61 ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTAGC  
 30 121 TCACTCATTA GGCACCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA  
 181 TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCTCA  
 241 GAATTAACCC TACTAAAGG GACTAGTCCT GCAGGTTTAA ACGAATTCGC CCTTAGCCAC  
 301 AGGCTTAATC TTCGGACATC CCCTTGAAAT GATTTCTAGG TACACAGCCC CCTGGCAGAG  
 361 CTGGGCTTCC CTGGACAGGG ACTGGCTGGA TTCCTGCTTC TGAAGATGGG ACTGTGGGCT  
 35 421 TGGTGGACAC TCCGGCCAGC CAGGACTGCT CCTGAGCAAC ACCGGGGGCA CCTGCTCCAG  
 481 GCCTTGATCT TGGGGAGGAA GCTAGGGTCT GAGGCTGCCT TGCCGAGGCC TCTGCACTGA  
 541 GCTTTTTTCAG GGCAGGGGAC TCCCACCCAT GTGGGCTCTC CCCAGGCCCT GCAGGAACGG  
 601 GCTGGCCTCC CGCACTGGAA CCACTGGGGT CGGGGCTCCT GGATTCCGCC GACAGTGCCT  
 661 CCAAAGCCG CTGCAGAGAT AGGGAGGCCG GGGCGGGGCT CCTACCAGTG CTGTGGGCAT

	721	GGGACCCCTCT	CCCACGGGGC	CTGGGGCCAC	TTGGGCTGCA	GCCGTAGAGT	TTGAGGACCC
	781	GGGTGATGTC	CGAGGCACTC	AGGTTCCATC	GCTGGCCGAT	GTGGACACTG	GGGGCCCAAA
	841	GTGGTGTGAT	GGTGGGCAGC	CCACGCCGGC	TGAAGGCGAG	CCTCCCATAG	TGCATCACAG
	901	AGGAGTAGTC	ATAGGGCGTC	AGCATGTTGC	TGCTCTGAGA	CTTGATGAAG	TTGATTTCAA
5	961	AGCCTGGCAG	GATCTCGTTC	CAGTTGACAC	GGATATAGCG	GTCCCGGTCTG	GCCCGCGTGT
	1021	GCTCGTGCCA	GAAGCCCAGC	ACATGCATGA	GCTCATGAAG	GACAAATGCCC	CGGCCCTTCT
	1081	GGAGACACGT	GGGCGCCAGG	GAGACCACCT	GCATCCCTCC	ACTGCGCCCC	ACATCTGAGA
	1141	AGCACCCATA	CATGGGGATG	ATGGAAATGA	AGTCTCTCTG	GTCTTGATAG	GTGACAAACC
	1201	TGATGCACGT	GGAACGTTCA	AACTCCGCAA	GAGCCTCCAG	GATGACCTGG	CGGCTGGGCT
10	1261	CATCGTACTT	GCTGGAGAGC	AGGAAGGGGA	CCTCCACGAC	ACCACTACCA	CCCATGGGCC
	1321	ATTTGTGTCT	GGTTGCTGAC	AGAAGGGCGA	ATTGCGGGCC	GCTAAATTCA	ATTCGCCCTA
	1381	TAGTGAGTCG	TATTACAATT	CACTGGCCGT	CGTTTTACAA	CGTCGTGACT	GGGAAAACCC
	1441	TGGCGTTACC	CAACTTAATC	GCCTTGACAG	ACATCCCCCT	TTGCCCAGCT	GGCGTAATAG
	1501	CGAAGAGGCC	CGCACCGATC	GCCCTTCCCA	ACAGTTGCGC	AGCCTATACG	TACGGCAGTT
15	1561	TAAGGTTTAC	ACCTATAAAA	GAGAGAGCCG	TTATCGTCTG	TTTGTGGATG	TACAGAGTGA
	1621	TATTATTGAC	ACGCCGGGGC	GACGGATGGT	GATCCCCCTG	GCCAGTGCAC	GTCTGCTGTC
	1681	AGATAAAGTC	TCCCGTGAAC	TTTACCCGGT	GGTGCATATC	GGGGATGAAA	GCTGGCGCAT
	1741	GATGACCACC	GATATGGCCA	GTGTGCCGGT	CTCCGTTATC	GGGGAAGAAG	TGGCTGATCT
	1801	CAGCCACCGC	GAAAATGACA	TCAAAAACGC	CATTAACCTG	ATGTTCTGGG	GAATATAAAT
20	1861	GTCAGGCATG	AGATTATCAA	AAAGGATCTT	CACCTAGATC	CTTTTCACGT	AGAAAGCCAG
	1921	TCCGCAGAAA	CGGTGCTGAC	CCCGGATGAA	TGTCAGCTAC	TGGGCTATCT	GGACAAGGGA
	1981	AAACGCAAGC	GCAAAGAGAA	AGCAGGTAGC	TTGCAGTGGG	CTTACATGGC	GATAGCTAGA
	2041	CTGGGCGGTT	TTATGGACAG	CAAGCGAACC	GGAATTGCCA	GCTGGGGCGC	CCTCTGGTAA
	2101	GGTTGGGAAG	CCCTGCAAAG	TAAACTGGAT	GGCTTTCTCG	CCGCCAAGGA	TCTGATGGCG
25	2161	CAGGGGATCA	AGCTCTGATC	AAGAGACAGG	ATGAGGATCG	TTTCGCATGA	TTGAACAAGA
	2221	TGGATTGCAC	GCAGGTTCTC	CGGCCGCTTG	GGTGGAGAGG	CTATTGCGCT	ATGACTGGGC
	2281	ACAACAGACA	ATCGGCTGCT	CTGATGCCGC	CGTGTTCGG	CTGTCAGCGC	AGGGGCGCCC
	2341	GGTCTTTT	GTCAAGACCG	ACCTGTCCGG	TGCCCTGAAT	GAAGTCAAG	ACGAGGCAGC
	2401	GCGGCTATCG	TGGCTGGCCA	CGACGGGCGT	TCCTTGCGCA	GCTGTGCTCG	ACGTTGTCAC
30	2461	TGAAGCGGGA	AGGGACTGGC	TGCTATTGGG	CGAAGTGCCG	GGGCAGGATC	TCCTGTCATC
	2521	TCACCTTGCT	CCTGCCGAGA	AAGTATCCAT	CATGGCTGAT	GCAATGCGGC	GGCTGCATAC
	2581	GCTTGATCCG	GCTACCTGCC	CATTGACCA	CCAAGCGAAA	CATCGCATCG	AGCGAGCACG
	2641	TACTCGGATG	GAAGCCGGTC	TTGTGATCA	GGATGATCTG	GACGAAGAGC	ATCAGGGGCT
	2701	CGCGCCAGCC	GAAGTGTTCG	CCAGGCTCAA	GGCGAGCATG	CCCGACGGCG	AGGATCTCGT
35	2761	CGTGACCCAT	GGCGATGCCT	GCTTGCCGAA	TATCATGGTG	GAAAATGGCC	GCTTTTCTGG
	2821	ATTCATCGAC	TGTGGCCGGC	TGGGTGTGGC	GGACCGCTAT	CAGGACATAG	CGTTGGCTAC
	2881	CCGTGATATT	GCTGAAGAGC	TTGGCGGCGA	ATGGGCTGAC	CGCTTCCTCG	TGCTTTACGG
	2941	TATCGCCGCT	CCCGATTGCG	AGCGCATCGC	CTTCTATCGC	CTTCTTGACG	AGTTCTTCTG
	3001	AATTATTAAC	GCTTACAATT	TCCTGATGCG	GTATTTTCTC	CTTACGCATC	TGTGCGGTAT
40	3061	TTCACACCGC	ATACAGGTGG	CACTTTTTCG	GGAAATGTGC	GCGGAACCCC	TATTTGTTTA
	3121	TTTTTCTAAA	TACATTCAA	TATGTATCCG	CTCATGAGAC	AATAACCCCTG	ATAAATGCTT



3181 CAATAATATT GAAAAAGGAA GAGTATGAGT ATTCAACATT TCCGTGTCGC CCTTATTCCC  
 3241 TTTTTTGC GG CATTTCGCCT TCCTGTTTTT GCTCACCCAG AAACGCTGGT GAAAGTAAAA  
 3301 GATGCTGAAG ATCAGTTGGG TGCACGAGTG GGTACATCG AACTGGATCT CAACAGCGGT  
 3361 AAGATCCTTG AGAGTTTTCG CCCC GAAGAA CGTTTTCCAA TGATGAGCAC TTTTAAAGTT  
 5 3421 CTGCTATGTG GCGCGGTATT ATCCCGTATT GACGCCGGGC AAGAGCAACT CGGTCGCCGC  
 3481 ATACACTATT CTCAGAATGA CTTGGTTGAG TACTCACCAG TCACAGAAAA GCATCTTACG  
 3541 GATGGCATGA CAGTAAGAGA ATTATGCAGT GCTGCCATAA CCATGAGTGA TAACACTGCG  
 3601 GCCAACTTAC TTCTGACAAC GATCGGAGGA CCGAAGGAGC TAACCGCTTT TTTGCACAAC  
 3661 ATGGGGGATC ATGTAACTCG CTTGATCGT TGGGAACCGG AGCTGAATGA AGCCATACCA  
 10 3721 AACGACGAGC GTGACACCAC GATGCCTGTA GCAATGGCAA CAACGTTGCG CAAACTATTA  
 3781 ACTGGCGAAC TACTTACTCT AGCTTCCCGG CAACAATTAA TAGACTGGAT GGAGGCGGAT  
 3841 AAAGTTGCAG GACCACTTCT GCGCTCGGCC CTTCCGGCTG GCTGGTTTAT TGCTGATAAA  
 3901 TCTGGAGCCG GTGAGCGTGG GTCTCGCGGT ATCATTGCAG CACTGGGGCC AGATGGTAAG  
 3961 CCCTCCCGTA TCGTAGTTAT CTACACGACG GGGAGTCAGG CAACTATGGA TGAACGAAAT  
 15 4021 AGACAGATCG CTGAGATAGG TGCCTCACTG ATTAAGCATT GGTAAGTGTG AGACCAAGTT  
 4081 TACTCATATA TACTTTAGAT TGATTTAAAA CTTTATTTT AATTTAAAAG GATCTAGGTG  
 4141 AAGATCCTTT TTGATAATCT CATGACCAA ATCCCTTAAC GTGAGTTTTC GTTCCACTGA  
 4201 GCGTCAGACC CCGTAGAAAA GATCAAAGGA TCTTCTTGAG ATCCTTTTTT TCTGCGCGTA  
 4261 ATCTGCTGCT TGCAAACAAA AAAACCACCG CTACCAGCGG TGGTTTGTTC GCCGGATCAA  
 20 4321 GAGCTACCAA CTC'TTTTTTCC GAAGGTAAGT GGCTTCAGCA GAGCGCAGAT ACCAAATACT  
 4381 GTCCTTCTAG TGTAGCCGTA GTTAGGCCAC CACTTCAAGA ACTCTGTAGC ACCGCCTACA  
 4441 TACCTCGCTC TGCTAATCCT GTTACCAGTG GCTGCTGCCA GTGGCGATAA GTCGTGTCTT  
 4501 ACCGGGTTGG ACTCAAGACG ATAGTTACCG GATAAGGCGC AGCGGTCGGG CTGAACGGGG  
 4561 GGTTTCGTGCA CACAGCCCAG CTTGGAGCGA ACGACCTACA CCGAACTGAG ATACCTACAG  
 25 4621 CGTGAGCTAT GAGAAAGCGC CACGCTTCCC GAAGGGAGAA AGGCGGACAG GTATCCGGTA  
 4681 AGCGGCAGGG TCGGAACAGG AGAGCGCACG AGGGAGCTTC CAGGGGAAA CGCCTGGTAT  
 4741 CTTTATAGTC CTGTCGGGTT TCGCCACCTC TGACTTGAGC GTCGATTTTT GTGATGCTCG  
 4801 TCAGGGGGGC GGAGCCTATG GAAAAACGCC AGCAACGCGG CCTTTTACG GTTCTGGGC  
 4861 TTTTGCTGGC CTTTGTGCTCA CATGTTCTTT CCTGCGTTAT CCCCTGATTC TGTGGATAAC  
 30 4921 CGTATTACCG CCTTTGAGTG AGCTGATACC GCTCGCCGCA GCCGAACGAC CGAGCGCAGC  
 4981 GAGTCAGTGA GCGAGGAAGC GGAAG

**SEQ ID NO: 39 (XpCR4TOPO IPAAA78836-2 plasmid nucleotide sequence)**

1 AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCATTAA TGCAGCTGGC  
 35 61 ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTAGC  
 121 TCACTCATTA GGCACCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA  
 181 TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCTCA  
 241 GAATTAACCC TACTAAAGG GACTAGTCCT GCAGGTTTAA ACGAATTTCG CCTTAGCCAC  
 301 AGGCTTAATC TTCGGACATC CCCTTGAAAT GATTTCTAGG TACACAGCCC CCTGGCAGAG  
 40 361 CTGGGCTTCC CTGGACAGGG ACTGGCTGGA TTCCTGCTTC TGAAGATGGG ACTGTGGGCT

421 TGGTGGACAC TCCGGCCAGC CAGGACTGCT CCTGAGCAAC ACCGGGGGCA CCTGCTCCAG  
481 GCCTTGATCT TGGGGAGGAA GCTAGGGTCT GAGGCTGCCT TGCCGAGGCC TCTGCACTGA  
541 GCTTTTTCAG GGCAGGGGAC TCCCACCCAT GTGGGCTCTC CCCAGGCCCT GCAGGAACGG  
601 GCTGGCCTCC CGCACTGGAA CCACTGGGGT CGGGGCTCCT GGATTCCGCC GACAGTGCCT  
5 661 CCAAAGCCG CTGCAGAGAT AGGGAGGCCG GAGCGGGGCT CCTACCAGTG CTGTGGGCAT  
721 GGGACCTCTT CCCACGGGGC CTGGGGCCAC TTGGGCTGCA GCCGTAGAGT TTGAGGACCC  
781 GGGTGATGTC CGAGGCACTC AGGTTCCATC GCTGGCCGAT GTGGACACTG GGGGCCCAAA  
841 GTGGTGTGAT GGTGGGCAGC CCACGCCGGC TGAAGGCGAG CCTCCCATAG TGCATCACAG  
901 AGGAGTAGTC ATAGGGCGTC AGCATGTTGC TGCTCCGAGA CTTGATGAAG TTGATTTCAA  
10 961 AGCCTGGCAG GATCTCGTTC CAGTTGACAC GGATATAGCG GTCCCGGTGCG GCGCGCTGT  
1021 GCTCGTGCCA GAAGCCCAGC ACATGCATGA GCTCATGAAG GACAATGCCC CGGCCCTTCT  
1081 GGAGACACGT GGGCGCCAGG GAGACCACCT GCATCCCTCC ACTGCGCCCC ACACTCGAGA  
1141 AGCACCCATA CATGGGGATG ATGGAAATGA AGTCTCTCTG GTCCTGATAG GTGACAAACC  
1201 TGATGCACGT GGAACGTTCA AACTCCGCAA GAGCCTCCAG GATGACCTGG CGGCTGGGCT  
15 1261 CATCGTACTT GCTGGAGAGC AGGAAGGGGA CCTCCACGAC ACCACTACCA CCCATGGGCC  
1321 ATTTGTTGCT GGTGCTGAC AGCAGTCGGA AGGGACTCGG CCGGATGATG TCCCCCTCGA  
1381 TGAGGAAGCT GCTCTCTGGG GTTCTTCCA GGATGAGCCC TTGGTTAATT GCAGGAATGT  
1441 CTTGTCCCC GGAGGCCTGG GTTCCCTCAG GGGTGAGGCC ATCTGGGAAG CTGGTACCAC  
1501 AGGCTCCTGC GCAGCTGGAG GCCAGGGGCG CTCTAGGAT CACACCTGGC AAGGAGAGCA  
20 1561 GACCCAGCAC CCAAGGCCAG AGACCCCCTA CACCCTCCAT GGTAAGAAAG GCGAATTCGC  
1621 GGCCGCTAAA TTCAATTCGC CCTATAGTGA GTCGTATTAC AATTCAGTGG CCGTCGTTTT  
1681 ACAACGTCGT GACTGGGAAA ACCCTGGCGT TACCCAACCT AATCGCCTTG CAGCACATCC  
1741 CCCTTTCGCC AGCTGGCGTA ATAGCGAAGA GGCCCGCACC GATCGCCCTT CCCAACAGTT  
1801 GCGCAGCCTA TACGTACGGC AGTTTAAGGT TTACACCTAT AAAAGAGAGA GCCGTTATCG  
25 1861 TCTGTTTGTG GATGTACAGA GTGATATTAT TGACACGCCG GGGCGACGGA TGGTGATCCC  
1921 CCTGGCCAGT GCACGTCTGC TGTCAGATAA AGTCTCCCGT GAACTTTACC CGGTGGTGCA  
1981 TATCGGGGAT GAAAGCTGGC GCATGATGAC CACCGATATG GCCAGTGTGC CGGTCTCCGT  
2041 TATCGGGGAA GAAGTGGCTG ATCTCAGCCA CCGCGAAAAT GACATCAAAA ACGCCATTAA  
2101 CCTGATGTTT TGGGGAATAT AAATGTCAGG CATGAGATTA TCAAAAAGGA TCTTCACCTA  
30 2161 GATCCTTTTC ACGTAGAAAG CCAGTCCGCA GAAACGGTGC TGACCCCGGA TGAATGTCAG  
2221 CTACTGGGCT ATCTGGACAA GGGAAAACGC AAGCGCAAAG AGAAAGCAGG TAGCTTGCAG  
2281 TGGGCTTACA TGGCGATAGC TAGACTGGGC GGTTTTATGG ACAGCAAGCG AACCAGGAAT  
2341 GCCAGCTGGG GCGCCCTCTG GTAAGGTTGG GAAGCCCTGC AAAGTAAACT GGATGGCTTT  
2401 CTCGCCGCA AGGATCTGAT GGCAGAGGG ATCAAGCTCT GATCAAGAGA CAGGATGAGG  
35 2461 ATCGTTTCGC ATGATTGAAC AAGATGGATT GCACGCAGGT TCTCCGGCCG CTTGGGTGGA  
2521 GAGGCTATTG GGCTATGACT GGGCACAACA GACAATCGGC TGCTCTGATG CCGCCGTGTT  
2581 CCGGCTGTCA GCGCAGGGGC GCCCGGTTCT TTTTGTCAAG ACCGACCTGT CCGGTGCCCT  
2641 GAATGAACTG CAAGACGAGG CAGCGCGGCT ATCGTGGCTG GCCACGACGG GCGTTCCTTG  
2701 CGCAGCTGTG CTCGACGTTG TCACTGAAGC GGGAAGGGAC TGGCTGCTAT TGGGCGAAGT  
40 2761 GCCGGGGCAG GATCTCCTGT CATCTCACCT TGCTCCTGCC GAGAAAGTAT CCATCATGGC  
2821 TGATGCAATG CGGCGGCTGC ATACGCTTGA TCCGGCTACC TGCCCATTCG ACCACCAAGC

2881 GAAACATCGC ATCGAGCGAG CACGTACTCG GATGGAAGCC GGTCTTGTCTG ATCAGGATGA  
 2941 TCTGGACGAA GAGCATCAGG GGCTCGCGCC AGCCGAAC TG TCGCCAGGC TCAAGGCGAG  
 3001 CATGCCCCGAC GGCGAGGATC TCGTCGTGAC CCATGGCGAT GCCTGCTTGC CGAATATCAT  
 3061 GGTGGAAAAT GGCCGCTTTT CTGGATTTCAT CGACTGTGGC CGGCTGGGTG TGGCGGACCG  
 5 3121 CTATCAGGAC ATAGCGTTGG CTACCCGTGA TATTGCTGAA GAGCTTGGCG GCGAATGGGC  
 3181 TGACCGCTTC CTCGTGCTTT ACGGTATCGC CGCTCCCGAT TCGCAGCGCA TCGCCTTCTA  
 3241 TCGCCTTCTT GACGAGTTCT TCTGAATTAT TAACGCTTAC AATTTCTCTGA TCGGGTATTT  
 3301 TCTCCTTACG CATCTGTGCG GTATTTTACA CCGCATACAG GTGGCACTTT TCGGGGAAAT  
 3361 GTGCGCGGAA CCCCTATTTG TTTATTTTTC TAAATACATT CAAATATGTA TCCGCTCATG  
 10 3421 AGACAATAAC CCTGATAAAT GCTTCAATAA TATTGAAAAA GGAAGAGTAT GAGTATTCAA  
 3481 CATTTCCGTG TCGCCCTTAT TCCCTTTTTT GCGGCATTTT GCCTTCTCTG TTTTGCTCAC  
 3541 CCAGAAACGC TGGTGAAAGT AAAAGATGCT GAAGATCAGT TGGGTGCACG AGTGGGTAC  
 3601 ATCGAACTGG ATCTCAACAG CGGTAAGATC CTTGAGAGTT TTCGCCCCGA AGAACGTTTT  
 3661 CCAATGATGA GCACTTTTAA AGTTCTGCTA TGTGGCGCGG TATTATCCCG TATTGACGCC  
 15 3721 GGGCAAGAGC AACTCGGTCT CGCATACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA  
 3781 CCAGTCACAG AAAAGCATCT TACGGATGGC ATGACAGTAA GAGAATTATG CAGTGTGCC  
 3841 ATAACCATGA GTGATAACAC TGCGGCCAAC TTAATTCTGA CAACGATCGG AGGACCGAAG  
 3901 GAGCTAACCG CTTTTTTGCA CAACATGGGG GATCATGTAA CTCGCCTTGA TCGTTGGGAA  
 3961 CCGGAGCTGA ATGAAGCCAT ACCAAACGAC GAGCGTGACA CCACGATGCC TG TAGCAATG  
 20 4021 GCAACAACGT TGCGCAAAC ATTAAC TGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA  
 4081 TTAATAGACT GGATGGAGGC GGATAAAGTT GCAGGACCAC TTCTGCGCTC GGCCCTTCCG  
 4141 GCTGGCTGGT TTATTGCTGA TAAATCTGGA GCCGGTGAGC GTGGGTCTCG CGGTATCATT  
 4201 GCAGCACTGG GGCCAGATGG TAAGCCCTCC CGTATCGTAG TTATCTACAC GACGGGGAGT  
 4261 CAGGCAACTA TGGATGAACG AAATAGACAG ATCGCTGAGA TAGGTGCCCTC ACTGATTAAG  
 25 4321 CATTGGTAAC TGTCAGACCA AGTTTACTCA TATATACTTT AGATTGATTT AAAACTTCAT  
 4381 TTTTAATTTA AAAGGATCTA GGTGAAGATC CTTTTTGATA ATCTCATGAC CAAAATCCCT  
 4441 TAACGTGAGT TTTTCGTTCCA CTGAGCGTCA GACCCCGTAG AAAAGATCAA AGGATCTTCT  
 4501 TGAGATCCTT TTTTCTGCG CGTAATCTGC TGCTTGCAAA CAAAAAACC ACCGCTACCA  
 4561 GCGGTGGTTT GTTTGCCGGA TCAAGAGCTA CCAACTCTTT TTCCGAAGGT AACTGGCTTC  
 30 4621 AGCAGAGCGC AGATACCAA TACTGTCCTT CTAGTGTAGC CGTAGTTAGG CCACCACTTC  
 4681 AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT  
 4741 GCCAGTGGCG ATAAGTCGTG TCTTACCGGG TTGGACTCAA GACGATAGTT ACCGGATAAG  
 4801 GCGCAGCGGT CGGGCTGAAC GGGGGGTTCTG TGCACACAGC CCAGCTTGGG GCGAACGACC  
 4861 TACACCGAAC TGAGATACCT ACAGCGTGAG CTATGAGAAA GCGCCACGCT TCCCGAAGGG  
 35 4921 AGAAAGGCGG ACAGGTATCC GGTAAGCGGC AGGGTCGGAA CAGGAGAGCG CACGAGGGAG  
 4981 CTTCCAGGGG GAAACGCCTG GTATCTTTAT AGTCTGTCTG GGTTCGCCA CCTCTGACTT  
 5041 GAGCGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC  
 5101 GCGGCCTTTT TACGGTTTCT GGGCTTTTGC TGGCCTTTTG CTCACATGTT CTTTCTGCG  
 5161 TTATCCCCTG ATTCTGTGGA TAACCGTAT ACCGCCTTTG AGTGAGCTGA TACCGCTCGC  
 40 5221 CGCAGCCGAA CGACCGAGCG CAGCGAGTCA GTGAGCGAGG AAGCGGAAG